SOCIAL COMPUTING, AUTUMN 2018: TERM PROJECTS

ID	Title	Brief description	Mentor	Students
SC 01	Diversification of Recommender Systems	Recommender systems use data on past user preferences to predict possible future likes and interests of different users. A key challenge is that while the most useful individual recommendations are to be found among diverse niche objects, the most reliably accurate results are obtained by methods that recommend objects based on user or object similarity. In this project we will try to address the challenge of diversifying the recommendation outputs without compromising too much on accuracy font.	Abhisek Dash assignmentad@gmail.com	1. Soham Poddar 18CS60R01 2. Shounak Paul 18CS60R02 3. Sangireddy Bharath Kum 11CS30030 4. M. Kalyan Bhargav 14CS10031
SC 03	Analysis and Classification of Drug-Users in Reddit	Online Social Media like Twitter, Reddit have become a great source of healthcare-related information.In this work our primary objective focuses on analyzing addiction/drug related subreddits to check whether they give sufficient information to comment on the state of addiction of the individual user. Based on this we will perform a user level classification. Broadly there will be 4 different classes ; addicted user, prone to addiction user, recovering user, recovered user.	Abhisek Dash assignmentad@gmail.com	1. Swastik Haldar 15CS10045 2. Atharva Vyas 15CS30005 3. Arkajyoti Pal 15CS30003 4. Shalmoli Ghosh 17CS72P07
SC 04	Bias and Segregation : A Multidimensional Opinion Dynamics Study on Polarization	After the birth of online media platforms, there has been a vast increase in the number of available viewpoints, perspectives, ideas and opinions, leading to a very diverse pool of information. So these platforms have been deploying many different kinds of algorithms, which decide on the flow of information reaching each of their target user. In most of these platforms, the algorithms try to optimize popularity and proximity of the information to the target while deciding on information for their target users. This is generally performed just to optimize their platform usage. However, critics have argued that these algorithms often introduce algorithmic bias, which in turn leads to polarization. In this study, we will represent each end user as an N-dimensional opinion vector, formally define algorithmic bias and segregation, and study the polarization in the online society using segregation.	Gourab Kumar Patro working.gkp@gmail.com	1. Parantap Dansana 15CS30046 2. Nilay Pochhi 15CS10033 3. Nishok Kumar Sundaressan 15CS30024 4. Sumeet Shirgure 15CS30035
SC 05	Building and Applying Legal Citation Network	Legal documents are typically hard to analyze because they are lengthy and there is extensive use of domain-specific terminologies. In this work we will look at this problem from a citation network perspective. We will build a legal citation network by collecting data from multiple sources where nodes will be Indian Supreme Court Cases and Acts present in the Indian legal system. We will then study the properties of this network and apply it to tasks like document similarity and prior case/precedent retrieval.	Abhisek Dash assignmentad@gmail.com	1. Paheli Bhattacharya 17CS92R05 2. Yashsvi Dixit 15CS10055 3. Vishwajeet Shukla 15CS10052 4. Subham Kotal 15CS30034
SC 06	A data driven approach of predicting vandalism in Wikipedia	In this project our goal is to characterise every edit action on a Wikipedia page into a set of factors which can be utilised to identify whether the edit action is a true edit or a anomalous edit.	Soumya Sarkar portkey1996@gmail.com	1. Pankaj Dhurve 15CS10031 2. Sayan Naskar 15CS30027 3. Vinay Singh 15CS30038 4. Ayush Sharma 15CS10008

SC 07	Understanding health complaints on Reddit	Online social media has increasingly being used by organizations or governments for sharing information and receiving feedback, complaints and reviews from the customers and citizens respectively. It is also being used for large-scale monitoring of both the physical and mental state of the users. Existing works have used Twitter, which is widely different from content on Reddit and other health discussion forums. In this term project, we first implement a complaint detection system based on Twitter (https://link.springer.com/chapter/10.1007/978-3-319-49586-6_44). We then extend the methodology to extracting health-related complaints from posts and reviews on Reddit.	Soumyadeep Roy soumyadeep.roy9@gmail.com	1. B Vishnu Vardhan 15CS10009 2. Raj Banoudha 15CS10035 3. Harsha Vattem 15CS10015 4. V Hemanth Reddy 15CS10051
SC 08	Understanding deep active learning with RNNs for text classification	We face the problem of scarcity of labeled data for running supervised learning algorithms, especially in the deep learning setting. Here, we try to understand how different active learning(AL) strategies perform with RNN, for the task of text classification. This will help when we only have very limited labeled data(~500-1000) available, like data labeled from Amazon Mechanical Turk. In this term project, we first implement this RNN paper (https://www.aaai.org/ocs/index.php/AAAI/AAAI15/paper/download/9745/9552). Then we experiment with different Active Learning strategies and compare their performance.	Soumyadeep Roy soumyadeep.roy9@gmail.com	1. Kumar Abhishek 15CS10022 2. Gautam Suman 15CS10014 3. Kunal Aggarwal 15CS10023 4. Satyam Sevanya 15CS10040
SC 09	Characterizing users in Location-based social networks through mining online check-in traces and network structure	Location-based social networks provide rich geo-tagged data about visits of online users to physical locations offline as well as comments of such users in the form of reviews. Can we mine this rich information along with the structure of the underlying social network to infer the behavior of users in LBSNs such as Yelp? We will leverage on publicly available data provided by Yelp dataset challenge for this project.	Ayan Kumar Bhowmick ayankumarbhowmick@gmail.com	1. Eeshan Bhaduri 18ID90J01 2. Arundhati Banerjee 14CS30043 3. Mousam Roy 14CS30019 4. Aarushi Agrawal 14EE35009
SC 10	Human emotion detection from typing activities performed in the smartphone	As a part of this project, we shall analyze the typing data collected from different users. This dataset is already available. It was collected by monitoring the typing activities of users as they performed different activities (like FB messenger, WhatsApp) on their smartphone. The dataset does not contain actual text, however contains different metadata like typing speed, typing mistake etc. (Details will be explained). We shall analyze the dataset to answer specific questions like (a) how accurately we can infer emotion using this data (b) as individual user's typing vary, can we propose an aggregate model to determine emotion (happy-unhappy, active-inactive) so that the need for personalized training is overcome (c) are there any specific group of emotion that are more easily detectable (d) can we use deep neural network (DNN) for emotion prediction?	Surjya Ghosh surjya.ghosh@gmail.com	1. Santosh T.Y.S.S 15CS30037 2. Bhanu Prakash Reddy 15CS10010 3. Aravind kollipaara 15CS10020 4. Mandepudi Manoj 11CS10027

SC 11	Active Zero Shot Learning on Social Media	Zero-shot learning aims at predicting a large number of unseen classes using only labeled data from a small set of classes and external knowledge about class relations.Given the ever-increasing amount of online contents, potentially hundreds of thousands of classes can be identified. However, it is not feasible to tag such a large amount of relevant contents for each class for training purposes. Instead, one has to actively decide which classes are the most useful to collect labeled data to train a zero-shot model to predict the remaining classes. We will address this issue mainly by analyzing and implementing a published paper in this area [https://link.springer.com/content/pdf/10.1007%2Fs41060-017-0042-5.pdf]	Anurag Roy anu15roy@gmail.com	1. Ayush Bansal 15CS30006 2. Achal Gupta 15CS30001 3. Tejas Gupta 15CS10046 4. Yash Jain 16MA20051
SC 13	Introspecting different Commercial Movie Recommendation Platforms	In this term project we will try to gather recommendation outputs from movie recommendation websites, e.g. Amazon Prime and Netflix, essentially to analyse different network centric patterns across the recommendation networks formed by these recommendation outputs.	Abhisek Dash assignmentad@gmail.com	1. Tarique Hussain 15IM10024 2. Anubhav Shukla 15IM30002 3. Krishnakant Deshmukh 15EE32001 4. Tushar Sinha 15IE10038
SC 14	Characterizing influential users in Twitter based on temporal retweet patterns	Finding influential users in a social network is a well studied problem. However, most existing works consider only network topological information to identify highly influential users. In this project, we want to leverage on temporal retweeting patterns of Twitter cascades to identify influential users that can capture heterogeneity in users' activity level in addition to network position. This project will involve data collection from Twitter to extract tweet information as well as follower network followed by processing this data to evaluate identified influential users.	Ayan Kumar Bhowmick ayankumarbhowmick@gmail.com	1. Veldurthy Sudhanyu 15CS10050 2. Addanki Surya Teja 15CS10002 3. Patibandla Harikrishna 15CS30022 4. K Rahul Dev 15CS10021